Working Papers

Modeling and Stochastic Simulation
of
NOM Reactions

Analysis supporting the project:
"Stochastic Synthesis - Simulating the Environmental Transformations
of Natural Organic Matter"

Steve Cabaniss
University of New Mexico

Summer 2002

Contents

1.)NOM Evolution: Defining the Problem
2.)Evolution of NOM: Starting Materials
3.)Reactions of NOM
4.)Calculating Reaction Probabilities

NOM Evolution: Defining the problem

The first step in constructing a model of any process is the problem definition. In this step, we think as scientists (rather than programmers) and ask "What do we want the model to achieve? What problems should it help solve?" More specifically, we need to ask what type of information will go into the model, what processes will be simulated, and what type of results we wish to produce. Once we have a good idea of what the model should do, we can begin to ask how it should do it- how to interpret our scientific model in the form of algorithms and data structures.

This document is intended to start our problem definition by posing a set of questions about the evolution of NOM. The list is not intended to be comprehensive, but rather a basis for further discussion.

1. What are the starting materials from which NOM arises? For example, must we consider lignin as a precursor molecule? Tannins? Proteins? Phospholipids?

2. What molecular attributes of NOM must be represented? Given that a true molecular representation (e.g., an *ab initio* molecular structure a la Gaussian, with calculated atomic positions and electron densities) is impractical, what will constitute an acceptable surrogate? Elemental composition? Functional group frequencies?'

3. What processes- geochemical, biological, hydrologic- must be modeled? Hydrolysis? Adsorption? Bacterial consumption? Downstream transport? Photo-oxidation?

4. What external environmental parameters must be included?

5. What functional and analytical properties of NOM must we be able to calculate or predict? While some properties like percent carbon or carboxyl acidity could be calculated trivially from molecular attributes, others (infrared and nmr spectra, polarity, redox potential, bio-availability, etc.) would require some special prediction calculations, which might be difficult to devise, time-consuming, or inaccurate. Which properties are essential to know, and which are merely convenient or desirable?

Since we must begin somewhere, I have attempted to answer these questions, at least in part. The purpose of these answers, as of the document, is to begin a discussion and not to conclude one. In that spirit, I anticipate many additions, emendations, and corrections from all of you, which can be incorporated in the next draft ....

1. Starting materials *What are the starting materials from which NOM arises? For example, must we consider lignin as a precursor molecule? Tannins? Proteins? Phospholipids?*

Historically, NOM molecules often have been hypothesized to derive from large biomolecules, especially lignins and tannins. It is also possible to think of small molecules, including some quite labile phosphoesters and metabolic intermediates, as being released during cell lysis and subsequently incorporated into NOM via condensation reactions. The list below primarily reflects the former approach, partly because the variety of small molecules is a daunting prospect.

Macromolecules vary to a great degree, of course- consider the proteome of a single-celled organism. To avoid dealing with each possible macromolecule, we can work with 'representative' molecules, which would have typical structures and elemental compositions. For example, a representative protein might have a molecular weight of 10,000, with an elemental composition reflecting the weighted average of the various amino acids. One problem is deciding which 'classes' of molecules are appropriate to represent as a single representative structure, and which must be treated distinctly; for example, should cellulose and hemi-cellulose be lumped together or treated separately? Another problem is finding the best way to determine the representative structure for a given class.

The following list strikes me as a minimum necessary to obtain reasonable similarity to the actual macromolecular precursors:

- polysaccharides
    - representative storage polysaccharide (starch?)
    - representative cellulose
- a representative protein
    - perhaps subdivided into 'pure' and glycoproteins?
- a representative polynucleotide
- a representative tannin
- a representative lignin structure
- a representative polyterpene
- a representative cutin

Smaller molecules which might also be important precursors:

- Phospholipids
- Sugars
- Amino acids
- Flavonoids
- Quinones

2. Molecular attributes *What molecular attributes of NOM must be represented? Given that a true molecular representation (e.g., an ab initio molecular structure a la Gaussian, with calculated atomic positions and electron densities) is impractical, what will constitute an acceptable surrogate? Elemental composition? Functional group frequencies?*

As noted in the proposal, we need an intermediate level of molecular representation which is more specific than 'percent carbon' but less detailed than a molecular connectivity map. Both elemental and functional group composition would appear to be essential. Here, the definition of 'functional groups' has been expanded to include rings, phenyl groups etc. Should it be further expanded to include specific 'biomarkers'?

Each molecule in the simulation would be represented as

A. Elemental composition
    Number of C, H, O, N, S and P atoms in the molecule

B. Functional group counts- numbers of
    Double-bond equivalents- calculable from the elemental data
    Total ring structures
    Phenyl groups

Alcohols	(alkyl C-O-H groups)
Phenols	(aromatic C-O-H groups)
Ethers	(C-O-C links)
Esters	(ester links, C(=O)-O-C)
Ketones	(ketone groups, C-C(=O)-C)
Aldehydes	(aldehyde groups, C-C(=O)-H)
Acids	(carboxylic acid groups, -COOH)
Aryl Acid	(COOH groups on aromatic rings)
Amines	(amine groups, C-NH2)
Ring N	(Nitrogen atoms in heterocyclic rings)
Amides	(amide links, C-C(=O)-N(-H)-C)
Thioethers	(thio-ether links, C-S-C)
Thiols	(thiol groups, C-S-H)
Phosphoesters	(phosphates bound to 3 C chains, R-O-P(=O)(OR)-OR)
H-phosphoesters	(phosphates bound to 2 C chains, R-O-P(=O)(OH)-OR)
Phosphates	(phosphates bound to 1 C chain, R-O-PO_3H_2)

C. Origin
    (The 'original' molecule is not well-defined after a bi-molecular condensation reaction, of course, but might still be useful in same cases.)
    Time molecule entered system
    Original type of molecule (protein, carbohydrate, etc.)
    Place of origin

3. Reactions and Processes *What processes- geochemical, biological, hydrologic- must be modeled? Hydrolysis? Adsorption? Bacterial consumption? Downstream transport? Photo-oxidation?*

I consider this to be the most complex and important section of the problem definition, and less well-developed than the sections on molecular attributes (2) and properties (5). While the list of reactions can expand as we proceed with the model, it would be helpful to begin with as broad a group of potential transformations as possible.

Physical reactions- (Note- these may be 'chemical' reactions in some sense, e.g. adsorption to a surface is a chemical phenomenon. The key distinction here is whether the molecular attributes are changed, or if only the position/state of molecule changes.)

- Adsorption to mineral phases
    - Initial adsorption
    - Migration to high-energy sites
    - Hemi-micelle formation at high coverage
- Aggregation
    - Metal-induced aggregation
- Formation of a micelle
- Transport downstream (surface water)
- Transport through porous media
- Volatilization

Chemical reactions- These reactions will change the molecular attributes.

- Abiotic bulk reactions
    - ester hydrolysis (Acid/base catalyzed)
    - amide hydrolysis (Acid/base catalyzed)
    - C=C hydration (Acid/base catalyzed)
    - ester condensation
    - amide formation
    - aldol condensation
    - thermal decarboxylation
    - metal-catalyzed oxidation of aromatic rings

Abiotic surface reactions

All of the above bulk phase reactions are expected to occur on surfaces as well. However, differing reaction conditions (surface pH, higher concentrations of molecules for condensations, greater availability of transition metal catalysts) will alter the reaction probabilities.

- Direct photochemical reactions
    - Photodecarboxylation
    - Metal catalyzed photo-oxidation

Indirect photochemical reactions

- OH radical addition
- HOOH oxidation
- O_2^- reduction

Extracellular enzyme reactions-

- Hydrolytic cleavage
- Oxidative cleavage
- Oxidations of aldehydes, phenols, quinones and alkenes
- Hydration
- Dehydration.

Microbial uptake of small molecules

(How many species or types of bacteria with differing uptake mechanisms should be postulated here?)

4. Environmental parameters *What external environmental parameters must be included?*

Initially, I think these parameters will probably be fixed at specific values (e.g., pH 7.5, 18 ºC), but some of them may reasonably be affected by NOM (light intensity, pH) and later model versions should reflect this potential 'feedback loop'.

Parameter	***Processes affected***
Temperature	All
pH	Most
Light intensity	Direct and indirect photo-reactions
Metal concentrations (Al(III) and Fe, especially)	Aggregation, micelle formation
Bacterial activity (how many types of bacteria must we postulate to provide a semblance of realism here?)	Microbial uptake
Water flow rate/pressure gradient	all transport
Surface area (how many surfaces?)	all surface reactions

5. Molecular Properties *What functional and analytical properties of NOM must we be able to calculate or predict? While some properties like percent carbon or carboxyl acidity could be calculated trivially from molecular attributes, others (infrared and nmr spectra, polarity, redox potential, bio-availability, etc.) would require some special prediction calculations, which might be difficult to devise, time-consuming, or inaccurate. Which properties are essential to know, and which are merely convenient or desirable?*

These properties are quantities which can be calculated from the elemental composition and functional group data. They may be useful and/or interesting because

a) they represent a specific molecular function of environmental interest (e.g., light absorbance)
b) they represent a molecular parameter which has been used as a predictor for an environmental function (13C-nmr aromaticity, which can be used to predict pollutant partitioning)
c) they represent a measurable quantity which could be used to calibrate/verify model performance (both light and absorbance and 13C-nmr aromaticity could be used in this way).

Some of the property calculations are trivial and almost exactly right- for example, molecular weight is the sum of all atomic weights. Others are both involved and necessarily approximate- for example, K_{ow} estimation is inexact even with complete structural knowledge, and requires extensive tables of 'fragment constants' to achieve even factor-of-two reliability. One of our tasks will be to devise efficient and reliable predictor functions for these properties.

Trivially calculated for each molecule

Molecular weight (MW)	e.g., #C x 12 + #H x 1 + #O x 16 +.....
Molecular charge (C)	molecular charge at pH 8 = #COOH - #amines
Charge density	C/MW

It would also be very easy to calculate particular elemental ratios, functional group ratios, etc.

Not-so-trivial

Note- some of these could be estimated pretty well using molecular modeling if we had an exact structure; the challenge is in devising estimation functions that work with the data available (the molecular attributes).

Bioavailability	fraction of carbon available for microbial uptake (I see this number used a lot, but I am not certain what it really means. Mike Perdue says he can estimate this fraction quite well from elemental composition data only. Suggestions? In particular, can we formulate a more mechanistic and less empirical approach to estimating bacterial uptake?)

Chemical Equilibrium constants- These may be difficult to predict. Certainly we cannot expect better than order-of-magnitude estimates, but even those might be useful if we can average over a large number of molecules.

K_{ow}	estimated octanol-water partition K. Chosen because it should be useful indicator of polarity and water solubility, and because we think we can measure it.
K_{Cu}	estimated Cu(II) binding constant. Chosen for comparison with voluminous published experimental data.
K_{ads}	estimated K of adsorption to goethite (iron oxide) Chosen for comparison with published data and because it will help predict adsorption within the model.
E^0	estimated standard potential for oxidation of the molecule (and concomitant reduction of a metal or other organic compound). Chosen for comparison with literature data on metal reduction.

UV-Visible Absorbance Spectra- Accurate spectral prediction will be difficult, but by averaging together many molecules we may be able to obtain reasonable overall results. One question- should the absorbance from each molecule be modeled as only one peak (almost certainly Gaussian) or as two? For each peak, we must know

λ_{max}	wavelength of maximum light absorbance
σ	standard deviation of Gaussian peak
ε	molar absorptivity at λ_{max}

Infra-red spectra- Using standard tables of functional group peak maxima, together with less extensive data on peak widths and intensities, we should be able to reproduce the gross features of many molecules. The spectra will tend to be dominated by the very intense peaks of carboxyl and -OH groups, while poor predictions of the fine structure due to the carbon skeleton vibrations will have only a minor impact on overall fit. Each functional group will have its own λ_{max} , σ, and ε.

^{13}C nmr spectra- It will be impossible to predict the splitting patterns and exact shifts without connectivity data, but we should be able to predict the gross ‘carbon type’ data frequently reported for NOM by grouping carbons by hybridization and functional group- carboxyls and esters in one group, aromatic C in a second group, alkyl chain (or cycle) C in a third, etc.

Evolution of NOM: Starting Materials

NOM is ultimately derived from the organic compounds in plants, either directly exuded or decomposed, or indirectly via animals. It therefore seems reasonable to begin a model of NOM evolution with biomolecules known to exist in plants. Wershaw (1994) has suggested that the basic classes of NOM precursors must include cellulose, hemicelluloses, lignin, lipids, cutins and proteins and amino acids. While each of these classes contains many different specific structures, it may be possible to devise 'average' structures which represent the structure and reactivity of the class. These 'average' structures can then be used as the starting materials in the stochastic NOM reactions.

Each 'average' structure must have the following properties:

Numbers of atoms for the elements C, H, N, O, S, and P
Numbers of functional groups and structural features:

Total ring structures	
Phenyl groups	
Alcohols	(alkyl C-O-H groups)
Phenols	(aromatic C-O-H groups)
Ethers	(C-O-C links)
Esters	(ester links, C(=O)-O-C)
Ketones	(ketone groups, C-C(=O)-C)
Aldehydes	(aldehyde groups, C-C(=O)-H)
Acids	(carboxylic acid groups, -COOH)
Aryl Acid	(COOH groups on aromatic rings)
Amines	(amine groups, C-NH2)
Ring N	(Nitrogen atoms in heterocyclic rings)
Amides	(amide links, C-C(=O)-N(-H)-C)
Thioethers	(thio-ether links, C-S-C)
Thiols	(thiol groups, C-S-H)
Phosphoesters	(phosphates bound to 3 C chains, R-O-P(=O)(OR)-OR)
H-phosphoesters	(phosphates bound to 2 C chains, R-O-P(=O)(OH)-OR)
Phosphates	(phosphates bound to 1 C chain, R-O-PO_3H_2)

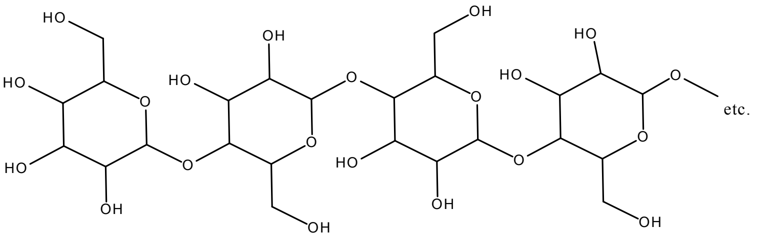
This information defines a molecule uniquely for the purposes of this simulation, although several different compounds might exist which could satisfy the conditions this information represents.

A. Cellulose

The average cellulose fragment described below represents a chain of 60 D-glucose units, linked together through carbons 1 and 4. Actual cellulose molecules are larger and insoluble; the average molecule described here is a smaller soluble fragment, presumably lysed from the parent molecule enzymatically. One question which arises is whether the O-linkages in this structure are adequately described as ether groups, or if they should be given a separate designation (1,1 di-ethers?) To indicate their special structure and reactivity.

Numbers of atoms

360	C
602	H
0	N
301	O
0	S
0	P



Numbers of functional groups and structural features:

60	Total ring structures
0	Phenyl groups
182	Alcohols
0	Phenols
119	Ethers
0	Esters
0	Ketones
0	Aldehydes
0	Acids
0	Aryl Acid
0	Amines
0	Ring N
0	Amides
0	Thioethers
0	Thiols
0	Phosphoesters
0	H-phosphoesters
0	Phosphates

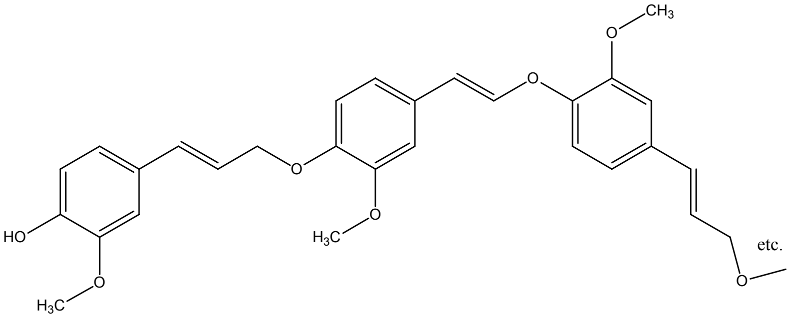
B. Hemicelluloses (to be added)

C. Lignin

The average lignin molecule described below represents an oligomer of 40 conferyl alcohol units condensed together via ether linkages (implying dehydration). This is a vast oversimplification of the variety of monomeric units in lignin, but retains the important features of high aromaticity and numerous aryl-alkyl ether linkages.

Numbers of atoms

400	C
322	H
0	N
81	O
0	S
0	P



Numbers of functional groups and structural features:

40	Total ring structures
40	Phenyl groups
1	Alcohols
1	Phenols
118	Ethers
0	Esters
0	Ketones
0	Aldehydes
0	Acids
0	Aryl Acid
0	Amines
0	Ring N
0	Amides
0	Thioethers
0	Thiols
0	Phosphoesters
0	H-phosphoesters
0	Phosphates

D. Lipids (to be added)

E. Cutins (to be added)

F. Proteins

Proteins occur in bewildering variety, but all are built upon a backbone of amino acids linked into a peptide chain. The average protein molecule described below is based upon the listed average amino acid content and a chain length of 50 residues.

Amino acid	R group
10% Glutamic acid	CH_2-CH_2-COOH
10% Lysine	CH_2-CH_2-CH_2-CH_2-NH_2
10% Glutamine	CH_2-CH_2-$CONH_2$
10% Serine	CH_2-OH
10% Threonine	CHOH-CH_3
10% Glycine	H
10% Alanine	CH_3
10% Valine	CH-$(CH_3)_2$
10% Leucine	CH_2-CH-$(CH_3)_2$
10% Phenylalanine	CH_2-C_6H_5

Numbers of atoms

240	C
332	H
60	N
76	O
0	S
0	P

Numbers of functional groups and structural features:

5	Total ring structures
5	Phenyl groups
10	Alcohols
0	Phenols
0	Ethers
0	Esters
0	Ketones
0	Aldehydes
6	Acids
0	Aryl Acid
6	Amines
0	Ring N
54	Amides
0	Thioethers
0	Thiols
0	Phosphoesters
0	H-phosphoesters
0	Phosphates

References

Ikan, R. 1991, Natural Products A Laboratory Guide Academic Press Inc. San Diego, California, 360 pp.

Lehninger, Biochemistry, 3rd edition

Wershaw, R.L. 1994, Membrane-Micelle Model for Humus in Soils and Sediments and its Relation to Humification USGS Water-Supply Paper 2410, US Government Printing Office, 48 pp.

Reactions of NOM

This document lists and attempts to briefly explain a subset of the reactions which we might expect NOM to undergo in the environment. Since it is written explicitly for the stochastic synthesis approach, it uses some of the terms and definitions from the proposed NOM structure and the reaction probability calculations described in other documents. In particular,

Δt = time interval considered for a possible reaction or movement (units of time)
P = probability of particular reaction (unitless)
k' = apparent first-order rate constant, which may represent more than one reaction pathway to the same results, with units of inverse time
In this document, we will assume all rate constants have units of hr^{-1} for convenient conversion into environmental timescales (as opposed to s^{-1}, which is common in the chemical literature)

Environmental parameters which affect the apparent rate constants include:

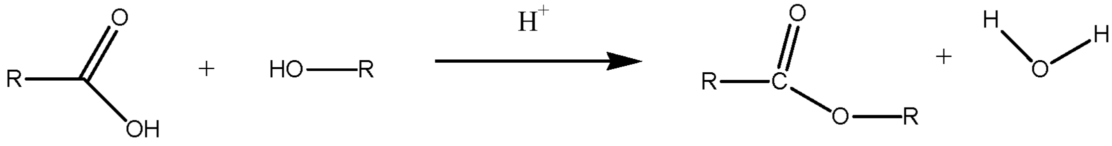
pH	the negative log of hydronium ion concentration
$[H^+]$	the concentration of hydronium ions in moles per liter, calculated from the pH as $[H^+] = 10^{-pH}$
$[OH^-]$	the concentration of hydroxide ions in moles per liter, calculated from the pH as $[OH^-] = 10^{pH-14.0}$
$[O_2]$	dissolved oxygen gas in moles per liter
MD	microbial density (What would be appropriate units for this? Do we need to distinguish bacteria from fungi, even at this early stage?)
I	light intensity in Einsteins per hour (1 mole of photons = 1 Einstein)
T	Temperature in Kelvins. 'Normal' temperature is 295 K (22 °C or 72 °F)

In general, the reaction probability is the product of k' and the time interval, $P = \Delta t\, k'$

This current reaction list is not comprehensive, or even a reasonable likeness of the final reaction list. However, we need a subset of reactions to test the software....

A. Ester condensation

This reaction represents the reaction of an alcohol with a carboxylic acid to create an ester linkage, thereby joining two molecules into one structure. This is the only second-order reaction in the current group of 10, and it would be expected to be quite slow in dilute solution due to the low probability of two NOM molecules encountering each other. For molecules in a micelle or adsorbed on a surface, the rate could be much higher, however, since NOM molecules would be in close proximity. In general, condensation occurs only if the acid is protonated- thus, the reaction is strongly pH dependent, and the rate is negligible at high pH values.



Transformation- All fields of the two molecules are summed to form a new molecule, which replaces one of the predecessor molecules (the other predecessor molecule disappears). The following changes are made to the molecular structure:

#O	-1
#H	-2
Acid	-1
Alcohol	-1
Ester	+1

The reaction probability for a given molecule is calculated

$P = \Delta t\ (\#\ \text{Acids})\ A\ e^{-Ea/RT} (1 + b\ [H^+])$ where $A =$

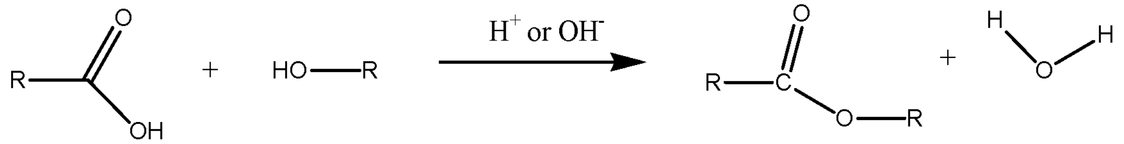
$E_a =$

$b =$

However, since this is a second order reaction, a second calculation must be made to determine if a second molecule with an alcohol group is 'close enough' to react. This calculation should depend on the number of molecules nearby. As an approximation, we could assume that for molecules in the micellar state, the probability is 1 that a nearby molecule exists (chosen randomly, then checked for presence of alcohol groups), while for adsorbed molecules or dissolved molecules the probability depends on surface concentration or dissolved concentration, respectively.

B. Ester hydrolysis

This is the reverse of reaction A. A molecule is split into two pieces by hydrolyzing an ester group to form an alcohol and a carboxylic acid. While this reaction is often quite slow at neutral pH, it can be accelerated by either acid **or** base catalysis, so that the rate is highly pH-dependent.



The transformation requires that the predecessor molecule be split into two successor molecules. First, one ester group is eliminated from the description of the predecessor, and one alcohol and one carboxylic acid are added. Next, a random number generator is used to obtain an 'overall' splitting ratio, R (say 50:50, or 70:30 or perhaps even 90:10). The new acid group is assigned to one successor, the new alcohol group to the other. Then, each remaining functional group is randomly assigned to either the larger or smaller successor molecule, along with necessary atoms (for example, a carboxylic acid group necessarily contains one C, two O and one H atom). Finally, the remaining atoms are assigned to a successor molecule with the stipulation that the number of H atoms in each successor be even unless the number of N atoms in that successor is odd.

#H	+2
#O	+1
Alcohol	+1
Acid	+1
Ester	-1

The reaction probability for a given molecule is calculated by

$P = \Delta t\,(\#\ \text{Esters})\,A\,e^{-Ea/RT}\,(1 + b\,[H^+] + c\,[OH^-])$ where $A =$

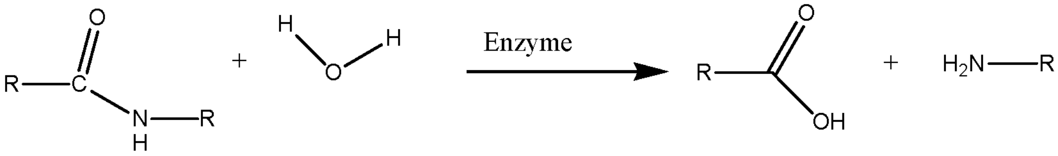
$E_a =$

$b =$

$c =$

C. Amide hydrolysis

This reaction splits a molecule at an amide linkage, adding a water molecule (2 H atoms and 1 O atom) while destroying the amide link and creating an amine and a carboxylic acid. This reaction is extremely slow at neutral pH, and requires extremely acidic/basic pH for abiotic catalysis. However, many bacteria and fungi synthesize and exude enzymes which catalyze the reaction effectively. The rate expression here relies upon these extracellular enzymes to promote amide hydrolysis.



The transformation requires that the predecessor molecule be split into two successor molecules. First, one amide link is eliminated from the description of the predecessor, and one amine and one acid are added. Next, a random number generator is used to obtain an 'overall' splitting ratio, R (say 50:50, or 70:30 or perhaps even 90:10). The new acid group is assigned to one successor, the new amine group to the other. Then, each remaining functional group is randomly assigned to either the larger or smaller successor molecule, along with necessary atoms (for example, a carboxylic acid group necessarily contains one C, two O and one H atom). Finally, the remaining atoms are assigned to a successor molecule with the stipulation that the number of H atoms in each successor be even unless the number of N atoms in that successor is odd.

#H	+2
#O	+1
Amine	+1
Acid	+1
Amide	-1

The reaction probability for a given molecule is calculated by

$P = \Delta t\ (\#\ \text{Amide})\ A\ e^{-Ea/RT}\ MD$ where $A =$

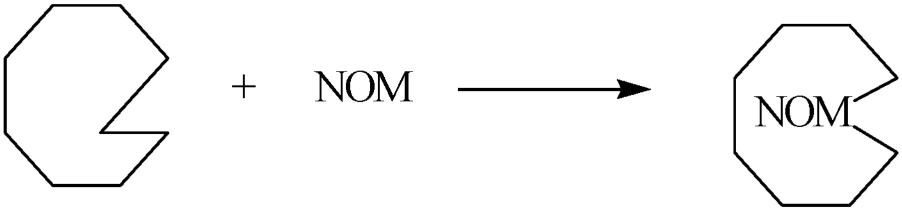
$E_a =$

(MD = microbial density)

D. Microbial uptake

This ‘reaction’ represents the likelihood that a micro-organism will envelop and assimilate/metabolize a given molecule. A MW constraint is in place to reflect the fact that molecules must be small enough for the microbe to envelop. The presence of N or P makes the molecule more ‘edible’ since these elements are typically in limited supply.

Small molecules (MW <100) have a very high probability of being consumed, and are consumed faster if they contain N or P atoms. Large molecules (MW > 1000) cannot be consumed, regardless of composition.



The transformation is simple- the molecule disappears from the simulation.

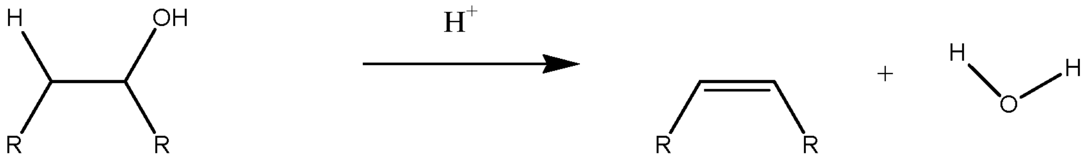
The reaction probability for a given molecule is calculated by

If $MW < 100$, $P = MD * (0.1 + (\#N + \#P) * 0.1)$
If $MW > 1000$, $P = 0$
If $100 < MW < 1000$, then
$P = MD * (0.1 - (0.0001\ MW) + 0.01\ (\#N + \#P))$

Note: MD represents microbial density, currently normalized to 1 for a dense, actively metabolizing community.

E. Dehydration

Dehydration represents the loss of an aliphatic alcohol group and an adjacent H atom to create a water molecule, and leaves a C=C bond on the organic molecule. This reaction usually proceeds by an acid-catalyzed pathway. At the moment, this reaction lumps together dehydration of primary, secondary and tertiary alcohols, even though the rates differ considerably (tertiary dehydrate fastest; primary, slowest).



The transformation of the molecule is simple, corresponding to the loss of one alcohol group, 2 H atoms and one O atom.

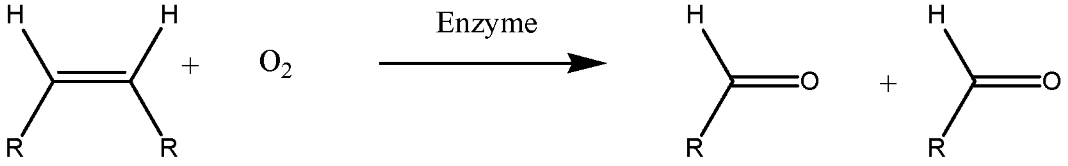
#H	-2
#O	-1
Alcohol	-1

The reaction probability for a given molecule is calculated by

$P = \Delta t\ (\#\ OH)\ A\ e^{-Ea/RT}\ (1 + b\ [H^+])$ where $A =$
$E_a =$
$b =$

F. Strong C=C oxidation

This reaction represents the oxidative cleavage of carbon-carbon double bonds (C=C). Reactions of this type are typically enzymatically mediated and can lead to a variety of products, depending on the location of the double bond. Photolytic oxidation can also occur. When the bond is located on a ring, ring-opening and loss of aromaticity occurs; otherwise, C=C cleavage leads to fragmentation of the original molecule. Furthermore, the products might be alcohols, aldehydes, ketones or acids depending on the substituents around the original C=C. Here, we assume that the cleavage creates two aldehydes, and the ratio of ring-opening to fragmentation is determined stochastically.



The transformation is

#O	+2	
Aldehyde	+2	
Phenyl		-1 (50%)

The transformation requires that the predecessor molecule be split into two successor molecules 50% of the time (or some other more realistic ratio). In the case of fragmentation, a random number generator is used to obtain an 'overall' splitting ratio, R (say 50:50, or 70:30 or perhaps even 90:10). One aldehyde group is assigned to each successor molecule. Then, each remaining functional group is randomly assigned to either the larger or smaller successor molecule, along with necessary atoms (for example, a carboxylic acid group necessarily contains one C, two O and one H atom). Finally, the remaining atoms are assigned to a successor molecule with the stipulation that the number of H atoms in each successor be even unless the number of N atoms in that successor is odd. In the case of ring-opening, the number of phenyl rings is decreased by one.

The reaction probability for a given molecule is calculated by

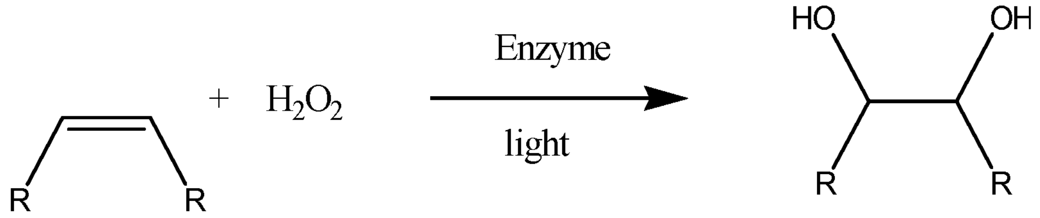
$P = \Delta t\ (\#C{=}C)\ A\ [O_2]\ (MD\ e^{-Ea/RT} + b\ I)$ where $A =$

$E_a =$

$b =$

G. Mild C=C oxidation

Mild C=C oxidation represents the oxidation of a carbon-carbon double bond (C=C) to form a pair of alcohols. No splitting of the molecule occurs, because the two carbons are still bonded (single bond only). This type of oxidation may occur via hydrogen peroxide (H_2O_2), which can be produced photochemically from dissolved O_2.



The transformation is

#H	+2
#O	+2

The reaction probability for a given molecule is calculated by

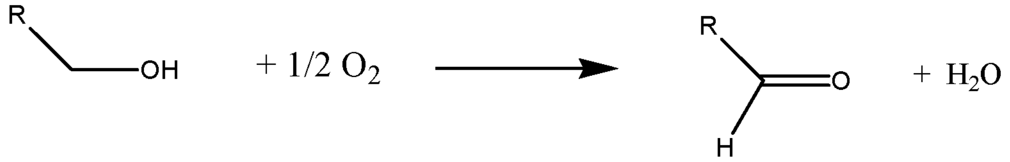
$P = \Delta t\ (\#C{=}C)\ A\ [O_2]\ (MD\ e^{-Ea/RT} + b\ I)$ where $A =$

$E_a =$

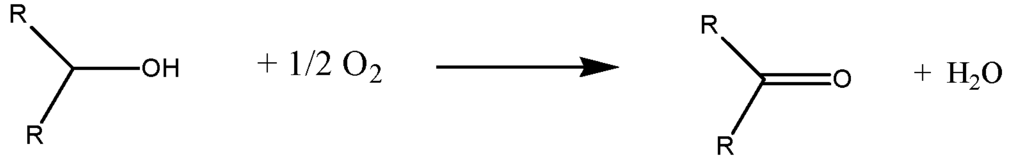
$b =$

H. Alcohol (C-O-H) oxidation

C-O-H oxidation represents the oxidation of a primary alcohol to form an aldehyde, or the oxidation of a secondary alcohol to form a ketone. The alcohol does not react directly with O_2, but with oxidized intermediates. However, since oxygen is the ultimate (albeit indirect) electron acceptor, the reactions can be written



OR



The overall transformation is simply the loss of two H atoms, with some probability (here 60%) of converting an alcohol into a aldehyde, otherwise (40%) into a ketone.

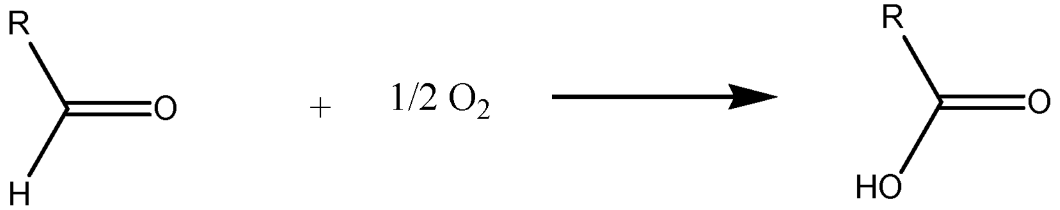
#H		-2
Alcohol		-1
60%	Aldehyde	+1
40%	Ketone	+1

The reaction probability for a given molecule is calculated by

$P = \Delta t\,(\#\,\mathrm{OH})\,A\,e^{-Ea/RT}\,[O_2]$ where A =
E_a =

I. Aldehyde C=O oxidation

C=O oxidation represents the oxidation of an aldehyde to a carboxylic acid by the insertion of a single O atom.



The transformation is

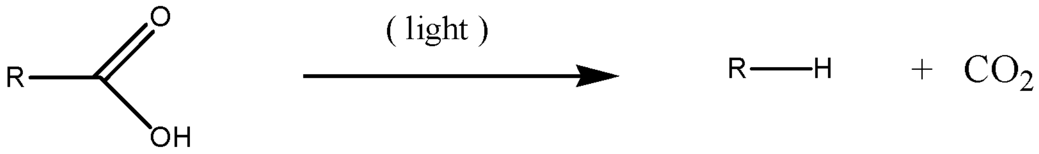
#O	+1
Aldehyde	-1
Acid	+1

The reaction probability for a given molecule is calculated

$P = \Delta t\ (\#\text{Aldehyde})\ A\ e^{-Ea/RT}\ [O_2]$ where $A =$
$E_a =$

J. Decarboxylation

Decarboxylation eliminates a carboxylic acid group as CO_2, leaving H behind. The reaction may proceed by a thermal or photochemical pathway, but in either case the rate is much higher for protonated carboxylic acid groups than for carboxylates. (The rate is greatly enhanced by the complexation of a redox-active metal, like Fe(III), but metal ions are not yet implemented in our model.)



The overall transformation is

-COOH	-1
#C	-1
#O	-2
Aryl carboxylic	-1 (60% of the time)

The reaction probability P for a given molecule is given by

$$P = \Delta t\ (\#\ \mathrm{COOH})\ A\ [H^+]\ (e^{-Ea/RT} + b\ I)$$

I = light intensity in Einsteins/hr
A =
E_a =
b =

Questions-

What is the probability of aliphatic versus aromatic decarboxylation?

Should the molar absorptivity of the molecule (or a surrogate like sp^2 %C) be included in the probability calculation?

Calculating Reaction Probabilities

A. The order of a reaction

The most commonly observed rate law governing macroscopic chemical kinetics is that for first order reaction (or first order decay) in which the rate or velocity of a reaction (R, units of concentration per time) depends only on the concentration of the reacting molecule (C units of concentration) and a rate constant k (units of reciprocal time)-

1. $R = k\,C$

The prevalence of this rate law results partly from the fact that many common reactions are truly ‘first order’ (i.e., the reaction rate is determined by reaction of that molecule only) and partly from the observation that many higher-order reactions (for example, second-order reactions in which the rate depends on the concentrations of two reacting molecules) show ‘apparent’ first order kinetics if the concentration of one reactant does not change significantly during the course of the reaction.

Apparent first order behavior can occur if one reactant is in vast excess (so that reacting a small portion of it makes a negligible change) or if the concentration of one reactant is buffered by other reactions (commonly observed for reactions with H^+ and OH^-, although not restricted to those reactants). In either case, a true second order reaction equation with two reactants (concentrations C_1 and C_2, rate constant k with units of reciprocal concentration and reciprocal time)

2. $R = k\,C_1\,C_2$

is reduced to an apparent first order equation

3. $R = k'\,C_1$

where the apparent first-order rate constant k’, also called the pseudo-first order constant, is given by

4. $k' = k\,C_2$

(assuming that C_2 is the reactant concentration that does not change significantly).

This is convenient, since first-order kinetics (true or apparent) are generally the simplest to treat stochastically.

B. Temperature dependence of rate constants: Arrhenius Theory

Temperature has a significant effect on the rates of chemical reactions. A ‘rule of thumb’ which applies to many simple reactions is that the rate of a reaction (and thus the k for that reaction) doubles for each 10 °C (or 10 K) increase in temperature. This ‘rule of thumb’ applies to many reactions of biological interest near room temperature, although it is easy to find counter-examples in the chemical literature. Probably the simplest and most general equation to relate k to T is the Arrhenius equation,

$$k = A\ e^{-Ea/RT}$$

where

A	is the pre-exponential factor (or collisional frequency factor) which is temperature independent and has the units of k
E_a	is the activation energy, commonly given in J or cal (or kJ or kcal) per mole of reaction
R	is the gas constant, 8.314 J mol^{-1} K^{-1}
T	is the temperature on the Kelvin scale

While it is probably unrealistic to assume that the A is independent of temperature, this equation has been used to describe the variation in observed k quite well over restricted temperature ranges. For purposes of these calculations, we will also assume that E_a is temperature independent for the sake of simplicity. (In any case, data on the temperature of E_a are not as commonly available as estimates of A and E_a.)

For many reactions, the pre-exponential factor is quite large- for second order reactions in aqueous media, diffusion calculations give $A \geq 10^9\ M^{-1}\ s^{-1}$. This would imply very large rate constants, except that the exponential term is typically quite small. For the ‘rule-of-thumb’ reaction discussed above, doubling the reaction rate as T increases from 293 K to 303 K corresponds to an activation energy $E_a = 51$ kJ/mole, which leads to an exponential term $e^{-Ea/RT}$ near 10^{-9} ! The overall k for such a reaction thus remains close to unity in spite of the magnitude of the pre-exponential term.

C. pH dependence of rate constants

Rates of aqueous reactions are often strongly pH dependent due to the participation of H^+ or OH^- or both as either a reactant or a catalyst. For example, the hydrolysis rates of esters may be catalyzed by either acid or base, leading to an apparent rate constant of the reaction

R-COO-R’ • •R-COOH + R’OH Rate = k’ [R-COO-R’]

which has a minimum near neutral pH with much higher values in acidic or basic solutions.

In some cases, the pH dependence arises from the protonated or deprotonated state of one or more reactants. These are referred to as pre-equilibrium states of reaction R, i.e., the proton transfer reactions are so fast that the reactant is assumed to reach protonation /deprotonation equilibrium before reaction R occurs. While not strictly true, this assumption is often very reasonable if the half-life of reaction R is on the order of milliseconds or more, since proton transfer in aqueous media reaches equilibrium in within microseconds. In these cases, the rate may be observed to depend on the fraction of reactant protonated (or deprotonated), leading to apparently complex rate expressions which depend on both the pH and the pK_a of the reactant.

More commonly, the pH dependence arises from direct participation of hydroxide or hydronium ions as a reactant or catalyst. In that case, the rate for a single mechanism is given by a simple rate law including the $[H^+]$ or $[OH^-]$ term. For example, the acid catalyzed ester hydrolysis rate for the above reaction is

$$R_{H+} = k\,[\text{R-COO-R'}]\,[H^+]$$

The overall rate law governing the reaction as a function of pH is the sum of all three mechanistic rate laws- un-catalyzed, acid-catalyzed and base-catalyzed- and can be expressed as an apparent first order rate constant k'

$$\begin{aligned} R_{total} &= R_{uc} + R_{H+} + R_{OH-} \\ &= k_{uc}\,[\text{R-COO-R'}] + k_{H+}\,[H^+]\,[\text{R-COO-R'}] + k_{OH-}\,[OH^-]\,[\text{R-COO-R'}] \\ &= (k_{uc} + k_{H+}\,[H^+] + k_{OH-}\,[OH^-])\,\,[\text{R-COO-R'}] \\ &= k'\,\,[\text{R-COO-R'}] \end{aligned}$$

where k' is a function of pH and the three mechanistic k values.

In this model we will use only this second type of pH dependence, which can be represented as an apparent first-order rate constant without requiring pK_a values.

D. Redox dependence of rate constants: Dissolved O_2 and electrochemical potential

Oxidation and reduction reactions of NOM are both important and complex. Numerous oxidizing species (O_2, Fe(III), H_2O_2, etc.) and reducing species (RS^-, O_2^-, Cu(I), etc.) are capable of reacting with organic compounds. Equilibrium calculations of redox speciation in the environment often rely upon the assumption of an equilibrium redox potential, E (or E_h in some texts) which represents the overall tendency for electron transfer. This is impractical for a kinetic model of NOM reactions, since

a) reactions with different oxidants (or reductants) proceed at different rates
and
b) reactions with different oxidants (or reductants) give different products

Unfortunately, modeling the concentrations of all possible oxidizing and reducing species in an environmental simulation would be difficult, if not impossible.

One possible compromise approach is to assume that in typical environments of interest to us, oxidation is more likely than reduction, and that since O_2 is the source of most other oxidants, oxidation reactions can be linked to the partial pressure of O_2 or to the concentration of dissolved O_2.

Now the rate of an oxidation reaction can be expressed as a function of dissolved O_2 and the organic molecule, leading to an apparent rate constant k' which depends upon $[O_2]$

$$R = k\,[O_2]\,[\text{organic}] = k'\,[\text{organic}]$$

For reactions in which the actual oxidant is not O_2, the relationship between $[O_2]$ and the concentration of the true oxidant (HOOH, Fe(III), etc.) must be assumed to be constant or at least calculable from known parameters (e.g., pH, total Fe, etc.).

E. Photochemical reactions: Dependence of rate on light intensity

Photochemical reactions can be split into two general categories:

1. Direct photoreactions require the reacting molecule to absorb a photon with sufficient energy to overcome the activation energy barrier. The rate of reaction is proportional not only to the light intensity, but also to the probability that the molecule will absorb a given photon (i.e., it is a function of the molar absorptivity of the reacting molecule).

2. Indirect photoreactions require that some other molecule (a photosensitizer) absorb a photon, creating a reactive intermediate which transfers energy to the reacting molecule. In this case the reacting molecule need not absorb light, but the sensitizer must. The rate of reaction is therefore proportional to the light intensity and to the

In general, photochemical reactions have a small, even negligible temperature dependence since typical ultraviolet and visible photons contain energy which greatly exceeds the thermal contribution. Thus the usual Arrhenius term can be replaced by a simple A term

F. Calculation of Reaction probabilities: Formalism

Consider a set of N reactions, each of which can occur with some probability P.

For the *i* th reaction, the first-order (or apparent first order) rate constant, k_i', has units of reaction events per unit time. Over a sufficiently small interval of time, the rate constant is thus the probability that a given molecule will react in that time, $P_{i,\Delta t}$.

$P_{i,\Delta t}$ = probability that a molecule will react via reaction *i* over a time interval Δt.
= $\Delta t\ k_i'$

Incorporating the effects of pH and/or dissolved O_2, this becomes

$$P_{i,\Delta t} = \Delta t\ k_i\ [R_i]$$

where $[R_i]$ is the constant concentration of H^+, OH^-, O_2, or other reagent in the ith reaction and k_i is the true rate constant for that reaction. Incorporating the effects of temperature, this becomes

$$P_{i,\Delta t} = \Delta t\ [R_i]\ A_i\ e^{-Eai/RT}$$

where A_i is the pre-exponential factor for reaction i and $E_{a\,i}$ is the activation energy for that reaction. For a photochemical reaction in which we assume no temperature dependence but instead a dependence on the light intensity I, the probability is given by

$$P_{i,\Delta t} = \Delta t\,[R_i]\,A_i\,I$$

In the case of parallel reaction mechanisms to reach the same products, the rate laws are additive. Thus, for a reaction which proceeds by both acid-catalyzed and uncatalyzed mechanisms, we expect a probability equation of the type

$$P_{i,\Delta t} = \Delta t\ \ A_i\,e^{-Eai/RT}\,(1 + b\,[H^+])$$

where b represents the multiplier necessary to change A_i into the rate constant for the catalyzed reaction pathway (this example assumes, unrealistically, that the E_a isthe same for both mechanisms).

G. References

Connors, K.A. 1990, Chemical Kinetics The Study of Reaction Rates in Solution VCH Publishers Inc., New York. 480 pp.

Espenson, J.H., 1981 Chemical Kinetics and Reaction Mechanisms McGraw-Hill, Inc. New York. 218 pp.

Erdi, P. And Toth, J. 1989, Mathematical Models of Chemical Reactions Princeton University Press, Princeton, NJ 259 pp.